DOCUMENT RESUME

ED 069 686 TM 002 137

AUTHOR Angoff, William H.
TITLE A Technique for the Investigation of Cultural Differences.
PUB DATE Sep 72
NOTE 11p.; Paper presented at American Psychological Association Meeting (Honolulu, Hawaii, Sept. 1972)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Analysis of Covariance; *Cluster Analysis; Comparative Analysis; *Correlation; *Cultural Differences; Factor Analysis; Interaction; *Mathematical Models; Measurement Techniques; *Research Methodology; Speeches; Statistical Analysis; Technical Reports
IDENTIFIERS Scattergram

ABSTRACT

A technique for detecting and studying item (or test) x group interactions independent of differences in level or dispersion of the groups is described. It involves construction of a scatter plot with two groups represented, one on each axis. Each point in the scatter plot represents the coordinates of a measure of a characteristic for one group plotted against a measure of the same characteristic for the other group; and the set of N points in the scatter plot represented the variables under study. The shape of the ellipse represents the degree to which the two groups in question share similar profiles. One method of analysis is to measure the departure of each point from the major axis of the ellipse and to study specific items that are most aberrant. Another is to calculate the variance associated with the item x interaction. Although the technique is not intended as a measure of item or test bias, it is useful in diagnosing cultural differences and comparing different types of groups. (Author/DJ)

A Technique for the Investigation of Cultural Differences

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

William H. Angoff
Educational Testing Service

APA Meetings
September 1972

ED 069686

The procedure I wish to describe here today for investigating cultural differences is by no means a new one, and although I have made use of it I am certainly not the first. Its first use, to my knowledge, was described by Thurstone in the 1920's in connection with his Method of Absolute Scaling. In this method indices of item difficulty -- i.e., p-values -- are obtained for two different groups on a number of items. Each p-value is converted to a normal deviate, and the pairs of normal deviates, one pair for each item, are plotted on a bivariate graph, each pair represented by a point on the graph. The plot of these points will ordinarily appear in the form of an ellipse extending from lower left to upper right, and if the two groups of individuals are drawn from the same type of population, the scatterplot of these points will fall on a long narrow ellipse, often representing a correlation of .98 or even higher, indicating that the rank order of difficulty of these items is essentially the same for the two groups. When the samples are somewhat different in level, the points will still fall in a long narrow ellipse, but displaced vertically or horizontally, depending on which group is the abler one. Even when the groups differ in dispersion, the points will still fall in the same type of ellipse, but the ellipse will be tilted at an angle more or less steeply than

TM 002137

ED 069686

TM 002 138

A TECHNIQUE FOR THE INVESTIGATION OF CULTURAL DIFFERENCES

by William H. Angoff
Educational Testing Service

This paper describes a technique for detecting and studying item (or test) x group interactions independent of differences in level or dispersion of the groups. Although the method is especially useful in the context of the study of cultural differences, it has been successfully applied in a variety of other contexts as well.

The method under consideration involves the construction of a scatter plot similar in appearance to the usual scatter plot of the correlation between two variables. In this method, however, unlike the usual scatter plot, there are two groups represented, one on each axis. Each point in the scatter plot represents the coordinates of a measure of a characteristic for one group plotted against a measure of the same characteristic for the other group; and the set of N points in the scatter plot represents the set of N characteristics, or variables, under study. Whereas the shape of the usual correlation ellipse represents the degree of association between the two variables in question, here the shape of the ellipse represents the degree to which the two groups in question share similar profiles on the set of variables under study. Thus, for example, a plot of the scale values of prestige for a series of occupations as they are perceived by one nationality group versus those perceived by another nationality group will help to reveal the nature of the differences in the perceptions of occupational prestige for the two nationalities and, by extension, will give some insights into the cultural differences of these two nationalities. Similarly, plots of item difficulties may be constructed for different regional or ethnic groups in the process of studying subtle differences in word meanings and uses.

Various methods of analysis have been attempted with data of this sort. One of them involves a measure of the departure of each point from the major axis of the ellipse and the study of the specific items that are most aberrant. Another makes use of more formal analytical techniques, involving the calculation of the variance associated with the item x group interaction. Although the general technique is not intended for use as a measure of item or test bias, it is useful in diagnosing cultural (and other group) differences and shows some promise for making equitable comparisons among different types of groups.

45°, depending on which group is the more dispersed. However, when the two groups differ in type, or when the items do not all have the same meaning for the two groups -- which may often be the case when the groups are drawn from the same general type of population but differ sharply in level or dispersion -- the item difficulties will not fall in precisely the same rank order for the two groups, and the correlation represented by these points will be lower than .98 or .99, sometimes substantially lower. The items falling at some distance from the plot may be regarded as contributing to the item x group interaction. They are the items that are especially more difficult for one group than for the other, relative to the other items, and they are the items that appear to represent a different "psychological meaning" to the members of the two groups.

Various investigators have chosen to examine the data of these bivariate plots in different ways. (It should be obvious that the plots are bivariate when there are two groups being compared at any one time. Presumably, however, any method of analysis designed for the comparison of two groups may be extended to three or more groups.) Cardall and Coffman (1964) describe the analysis of the item x population -- or item x race or item x culture -- interaction, in which random subgroups are nested within the larger categories of population, race, or culture. Cleary and Hilton (1968) followed the same general design except for the fact that they chose to nest three levels of SES within race. It is interesting to note that although the item x race variance was highly significant in their study,

it accounted for less than two percent of the total variance.

Gulliksen and Tucker have also used the bivariate plot in studies of cultural differences. Gulliksen (1960) derived scale values of occupational prestige based on responses by Frenchmen and also by Belgians for each of 31 occupations, and plotted the pairs of scale values for each of the occupations. The aberrant points -- those that fell away from the elliptical swarm of points -- represented occupations that were differently regarded by these nationalities. It is interesting that the most aberrant of the 31 occupations are religious occupations: missionary and clergyman. These are relatively highly regarded by the Belgians, but relatively poorly regarded by the Frenchmen.

Tucker, in an unpublished paper in 1953 described the analysis of a set of vocabulary items administered to students in Texas and in the Northeastern U. S. Generally, the items fell along the typical narrow elliptical pattern, with the exception of one or two dramatic aberrations. "Scorpion" was a word that was easier, relative to the other items, for Northerners than for Texans, and this despite the fact that the animal is indigenous to Texas, and not to the Northeast. A little sleuthing uncovered the fact that the Texas students, who knew the animal well, knew it as "stinglizard," not as "scorpion."

I have also had occasion to use the bivariate plot in a variety of contexts. In an effort to determine whether the Law School Admission Test was biased against Canadian students, Christina Herring and I (1971) collected two non-random samples of Americans and one sample of Canadians, and plotted the item difficulties, as described above, for Canadians vs. each of the American groups, and also for one of the American groups vs. the other. The result of this work was to show that the pattern of points for the two different national groups was about .98, not noticeably different from the pattern of points for one of the American groups vs. the other. The Canadian group was behaving no different with respect to their perception of these items than the American group, and there was therefore no reason to believe that the test was discriminating unfairly against either group relative to the other.

In another investigation Amiel Sharon and I (1971) compared the performance of a group of American students on the Test of English as a Foreign Language (TOEFL) with the performance on that test of foreign applicants to American universities. The intent here was to determine whether the test was capitalizing on the kinds of errors in English typically made by foreigners, or whether it was discriminating -- to no purpose -- among American students. As expected, the plots of item difficulty values for the different parts of the test showed very low correlations, ranging from .16 to .74, attesting to the fact that people with entirely different

orientations to these items were being studied. An editorial examination of the items that were especially difficult for the Americans then revealed that these were items calling for relatively formal and obscure discriminations in the English language, the kinds of discriminations that American students are not ordinarily trained to make.

This study triggered off another one in which the item responses to TOEFL were compared for six different language groups (1972). In this study the distance of each point in the bivariate ellipse from the major axis of the ellipse was determined as a measure of the item x group interaction of that item. The intent was to determine whether items with extreme distance-values could be characterized in some way that would permit us to make some generalizations regarding the kinds of errors in English characteristically made by each language group.

In another study Susan Ford and I attempted to make further investigations of the item x group interaction for Blacks and Whites. First, we made plots of item difficulties for samples between these races and between random samples within races, and discovered, as we expected, that there was a clear item x race interaction; the between-race plots represented correlations that were lower than those within races. Next, hypothesizing that a large part of this interaction was attributable to the simple differences between Blacks and Whites in level of performance, we matched the groups on an external measure, calculated item difficulties again and plotted again. The hypothesis was supported in that the correlation between the

indices for the two races rose, not quite to the level of a correlation between random groups, but to a clearly higher level nevertheless. And very likely it would have risen still higher had we used a set of matching variables that were more highly correlated with the performance variables under explicit study than the one we actually did use.

Item plots were also made for Blacks living in different urban areas (Atlanta and Savannah), and still other plots were made for Blacks living in urban areas vs. Blacks living in rural areas. Although these plots did not lead to conclusive results, they did suggest the direction of further research.

At the present time Christopher Modu and I are preparing a report of an attempt to develop a conversion of scores across tests designed for groups of different languages and cultures. This is the conversion of scores on the verbal and mathematical sections of the College Board Prueba de Aptitud Academica, designed for Puerto Rican students and expressed in Spanish, to scores on the corresponding verbal and mathematical sections of the College Board Scholastic Aptitude Test, which, as you know, is designed for U. S. students and expressed in English. The method we adopted in this study was to assemble groups of items in English and in Spanish, to translate each item into the other language and to administer all the items -- 155 verbal items and 100 mathematical items -- in the appropriate language mode to both

language groups, those taking the PAA in Spanish and those taking the SAT in English. As in the other studies just described we made plots of item difficulties, one plot for the verbal items and another for the mathematical items, in which one axis for these plots represented the scale of item difficulties for the PAA group and the other, the corresponding scale of item difficulties for the SAT group. As one would expect, the plots were highly dispersed, representing a correlation of only .60 for the verbal items and .87 for the mathematical items, clearly indicating a strong item x group interaction. Although some of this interaction results from the difficulty in making adequate translations, it should also be pointed out that this very difficulty is a function of what we mean by item x group interaction, or in a non-statistical sense, a function of the very nature of the cultural difference we were attempting to bridge here.

From these 155 verbal and 100 mathematical items we chose 40 verbal and 25 mathematical items principally on the basis of their distance from the major axis of the correlation ellipse, giving preference, of course, to those that were closest to that line. Other factors also weighed heavily in the choice: item difficulty, item discrimination, and whenever possible, considerations relating to a reasonable representation of the item types that normally appear in the operational Spanish and English tests.

Once these items were chosen -- items which were thought to contribute relatively little item x group interaction and therefore to represent very nearly the same "psychological meaning" for the two language groups -- they were given again as equating items in the appropriate language mode along

with the operational forms of the Spanish and English tests at regular administrations of those tests. The data on these "common items" were used to calibrate for differences between the Puerto Rican and U. S. students, permitting the equating of Spanish-language and English-language SATs.

The item plots described here have been, and I hope will continue to be, used in a variety of contexts and purposes to lead the way to the formation of hypotheses about group differences that go beyond the simple differences in means and standard deviations. The special advantage of these plots is that they can demonstrate quite dramatically both graphically and analytically the general presence of an item x group interaction and can help to identify the specific items or variables that are contributing most heavily to that interaction. In this way I am quite sure that they can prove to be a valuable tool in a wide variety of studies, in particular, in studies of differences between subcultures that coexist within this country and quite possibly between the cultures of geographically separate countries.

WHA:as

References

Angoff, W. H. and Ford, Susan F. Item-race interaction on a test of scholastic aptitude. Research Bulletin 71-59; CEEB RDR 71-72, No. 3. Princeton, N. J.: Educational Testing Service, 1971.

Angoff, W. H. and Herring, Christina L. Study of the appropriateness of the Law School Admission Test for Canadian and American students. Princeton, N. J.: Educational Testing Service, 1971. Unpublished manuscript.

Angoff, W. H. and Sharon, A. L. A comparison of scores earned on the Test of English as a Foreign Language by native American college students and foreign applicants to U. S. colleges. TESOL Quarterly, 1971, 5, 129-136.

Angoff, W. H. and Sharon, A. L. Patterns of test and item difficulty for six foreign language groups on the Test of English as a Foreign Language. Research Bulletin 72-2; CEEB RDR 71-72, No. 5. Princeton, N. J.: Educational Testing Service, 1972.

Cardall, Carolyn and Coffman, W. E. A method for comparing the performance of different groups on the itemx in a test. Research Bulletin 64-61, Princeton, N. J.: Educational Testing Service, 1964.

Cleary, T. Anne and Hilton, T. L. An investigation of item bias. Educational and Psychological Measurement, 1968, 28, 61-75.

Gulliksen, H. Intercultural attitude comparisons and introductory remarks at the Princeton University Conference on Preference Analysis and Subjective Measurement Research Memorandum 60-8, Princeton, N. J.: Educational Testing Service, 1964.